Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 95 (2016) 254 – 263

Procedia Computer Science

Complex Adaptive Systems, Publication 6
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2016 - Los Angeles, CA

Individual Compositional Cluster Analysis

Mika Sato-Ilic*

Faculty of Engineering, Information and Systems, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan

Abstract

This paper proposes an Individual Compositional Cluster Analysis with the aids of conventional individual difference scaling (INDSCAL) model. Our target data is 3-way data which consists of objects, variables, and subjects and we treat the case when the number of objects is relatively large. The purpose of this method is to show the difference of individual subjects visually in a lower dimensional space including the feature of objects with respect to variables as clusters. Numerical examples using sensor data obtained from several subjects show a better performance.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology

Keywords: Fuzzy clustering; compositional data; 3-way data; time-series data; Individual Difference Scaling; fault diagnosis

1. Introduction

Recently, data analysis for large multiple source time-series data, such as sensor information or server log data, has attracted interest as the need for real-time accumulated data has grown. For example, if we observe values of sensor data to measure individual human body activity and the data is consisted of observed times with respect to several kinds of sensors over multiple numbers of persons, then this data can be treated as a 3-way data in which

* Corresponding author. Tel.: +81-29-853-5006.
E-mail address: mika@risk.tsukuba.ac.jp

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2016.09.330

objects are times, variables are several kinds of sensors, subjects are the multiple numbers of persons. From this data, if we know the difference among subjects through features of clusters of times with respect to several kinds of sensors, then this will be useful to diagnose an individual subject's body activity. If these subjects are machines, then known difference among several machines over the times with respect to several kinds of sensors can contribute for fault diagnosis of the machines. In addition, if we can see the difference of subjects visually in a lower dimensional space, then this will be useful. Therefore, this paper proposes a method named "Individual Compositional Cluster Analysis" to show difference among several subjects visually in a lower dimensional space when we observe a 3-way data consisted of objects, variables, and subjects.

Conventionally, INDSCAL[1,2] model can show the difference among several subjects visually in a lower dimensional space when we observe a 3-way data. However, if the observed number of objects is large, then the lost of original data information by using the INDSCAL model is large and we can not safficientlty obtain a good fitness to the original data.

In order to solve this problem, the Individual Compositional Cluster Analysis exploits a result of fuzzy clustering for 3-way data. That is, if the clusters can summerize the objects with respect to variables and if we can obtain exactly the "same" clusters over the subjects, then we can use the obtained clusters as features representing objects, replaced by original objects. In order to obtain exactly the same clusters over the subjects, we propose a fuzzy clustering method for 3-way data in the Individual Compositional Cluster Analysis. Since the number of clusters is much smaller than the number of objects, we can fit the INDSCAL model well for the clusters. In addition, in order to locate the obtained clusters in a lower dimensional Euclidean vector space, we need to provide "common" dimensions over the subjects to the obtained clusters by using the "common" scale over the subjects. To provide these common dimensions to the obtained clusters over the subjects, we propose a transformation of combination between a result of a fuzzy clustering for a 3-way data and a result of INDSCAL in the Individual Compositional Cluster Analysis. The transformation is essentially similar to our previously proposed transformation[3], but also, different from the previous transformation at a point, that is, in the case of the Individual Compositional Cluster Analysis, the clustering result obtained at an individual subject over the same clusters and the result of INDSCAL obtained as single result over the subjects are used.

This paper consists of the following: Section 2 describes conventional INDSCAL model. Section 3 proposes a fuzzy clustering for 3-way data. Section 4 proposes the Individual Compositional Cluster Analysis. Section 5 proposes a criterion for the selection of an adaptable number of clusters for the Individual Compositional Cluster Analysis. Section 6 presents numerical examples of the Individual Compositional Cluster Analysis for a 3-way sensor data and Section 7 contains the conclusion. So, Sections 3 to 6 detail the original study of this paper.

2. Individual Difference Scaling

Multidimensional scaling (MDS)[4,5,6] is a method for capturing efficient information from observed dissimilarity data by representing the data structure in lower dimensional spatial space. Suppose the observed dissimilarity among n objects as follows:

$$D = (d_{ij}), \quad i, j = 1, \cdots, n, \; d_{ij} = d_{ji}, (i \neq j). \tag{1}$$

As a metric MDS, the following model[4,5] has been proposed.

$$d_{ij} = \{\sum_{\lambda=1}^{R} (x_{i\lambda} - x_{j\lambda})^2\}^{\frac{1}{2}} + \varepsilon_{ij}, \quad i, j = 1, \cdots, n. \tag{2}$$

In (2), d_{ij} is an observed dissimilarity between objects i and j shown in (1) and $x_{i\lambda}$ is a point of an object i with respect to dimension λ in R dimensional configuration space and $R < n$. ε_{ij} is an error. That is, MDS finds R dimensional scaling (coordinate) $(x_{i1}, \cdots, x_{iR})$ and throws light on the structure of the similarity relationship among the objects by representing the observed d_{ij} as the distance between a point $(x_{i1}, \cdots, x_{iR})$ and a point $(x_{j1}, \cdots, x_{jR})$ in R dimensional space. If the dissimilarity among n objects shown in (1) is observed by T subjects, then the data is represented as follows:

$$D_t = (d_{ij,t}), \quad i, j = 1, \cdots, n, \; t = 1, \cdots, T, \; d_{ij,t} = d_{ji,t}, (i \neq j). \tag{3}$$

Individual difference scaling (INDSCAL)[1,2] is a method for capturing efficient information from observed T dissimilarity data by representing the T data structures in the common lower dimensional spatial space. By using the common lower dimensional space as the common scale over the T subjects, the individual difference among the T subjects is represented by the individual weights for the obtained dimensions spanning the common lower dimensional space. The model of the INDSCAL has been proposed as follows:

$$d_{ij,t} = \{\sum_{\lambda=1}^{R}(\sqrt{w_{t\lambda}}x_{i\lambda} - \sqrt{w_{t\lambda}}x_{j\lambda})^2\}^{\frac{1}{2}} + \varepsilon_{ij,t} = \{\sum_{\lambda=1}^{R} w_{t\lambda}(x_{i\lambda} - x_{j\lambda})^2\}^{\frac{1}{2}} + \varepsilon_{ij,t}, \quad i, j = 1, \cdots, n, \; t = 1, \cdots, T. \tag{4}$$

In (4), $d_{ij,t}$ is an observed dissimilarity between objects i and j at t-th subject shown in (3) and $w_{t\lambda}$ is a weight of a subject t with respect to dimension λ in R dimensional configuration space and $R < n$. $\varepsilon_{ij,t}$ is an error at t-th subject. That is, INDSCAL finds R dimensional scaling (coordinate) $(x_{i1}, \cdots, x_{iR})$ over the T subjects commonly and the weights $w_{t\lambda}$ which show individual salience for each common dimension. Therefore,

$$o_{i\lambda,t} \equiv \sqrt{w_{t\lambda}}x_{i\lambda}, \; t = 1, \cdots, T \tag{5}$$

in (4) shows individual scale at each t subject, however, this scale is comparable over the T subjects, since the dimension λ is exactly the same dimension over the T subjects. Suppose the observed dissimilarity $d_{ij,t}$ is a Euclidean distance in (4), then $\varepsilon_{ij,t} = 0, \forall i, j$ and (4) can be rewritten as follows:

$$D_t^2 = \mathbf{11}' diag(XW_tX') - 2(XW_tX') + diag(XW_tX')\mathbf{11}', \tag{6}$$

where $D_t^2 = (d_{ij,t}^2)$, $\mathbf{1} = (1, \cdots, 1)'$, $X = (x_{ij})$, $i, j = 1, \cdots, n$, and W_t is a diagonal matrix consisted of diagonal elements $w_{t1}, \cdots, w_{tn}$. $diag(A)$ means a diagonal matrix which consists of diagonal elements of A. From (6) and the Young-Householder transformation[6], D_t^2 in (6) can be transformed as follows:

$$P_t = -\frac{1}{2}JD_t^2J = JXW_tX'J = XW_tX', \quad JX \equiv X, \tag{7}$$

where $P_t = (p_{ij,t})$, $i, j = 1, \cdots, n$, $t = 1, \cdots, T$. Matrix J is a symmetric matrix whose diagonal elements are $1 - 1/n$ and non-diagonal elements are $-1/n$ which means the centering operation for each column of X, that is the following condition

$$\sum_{i=1}^{n} x_{i\lambda} = 0, \quad \forall \lambda \tag{8}$$

is satisfied in order to fix an origin as 0 for all n dimensions in the obtained coordinate space. If the following condition $\frac{1}{T}\sum_{t=1}^{T} W_t = I$ is assumed for the weights, where I is a unit matrix, then

$$\overline{P} = \frac{1}{T}\sum_{t=1}^{T} P_t = \frac{1}{T}\sum_{t=1}^{T} XW_tX' = X\left(\frac{1}{T}\sum_{t=1}^{T} W_t\right)X' = XX'. \tag{9}$$

By using eigenvalue decomposition of $\overline{P}$ in (9), $\overline{P}$ can be represented as follows:

$$\overline{P} = V\Delta V', \tag{10}$$

where $V = (v_{ij}) = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$, $\mathbf{v}_\lambda = (v_{1\lambda}, \cdots, v_{n\lambda})'$, $i, j = 1, \cdots, n$, and Δ is a diagonal matrix whose diagonal elements are eigen values $\delta_1, \cdots, \delta_n$ and satisfy $\delta_1 > \cdots > \delta_n$. V is a matrix whose column vectors are eigen vectors $\mathbf{v}_1, \cdots, \mathbf{v}_n$ corresponding to eigen values $\delta_1, \cdots, \delta_n$. From (10), we obtain as follows:

$$\overline{P} = V\Delta^{\frac{1}{2}}\Delta^{\frac{1}{2}}V' = (V\Delta^{\frac{1}{2}})(V\Delta^{\frac{1}{2}})', \tag{11}$$

where $\Delta^{\frac{1}{2}}$ is a diagonal matrix whose diagonal elements are eigen values $\sqrt{\delta_1}, \cdots, \sqrt{\delta_n}$. From (9) and (11), if we assume that one solution of X is Y, then Y can be obtained as follows:

$$Y = V\Delta^{\frac{1}{2}}. \tag{12}$$

Then from (9), for any $n \times n$ orthogonal matrix L, X has to satisfy the following equation.

$$X = YL. \tag{13}$$

In order to determine L, the following linear combination is defined, where $f_1, \cdots, f_T$ are any scalars.

$$\tilde{P} = \sum_{t=1}^{T} P_t f_t = X\left(\sum_{t=1}^{T} W_t f_t\right)X' = X\tilde{W}X', \ \tilde{W} \equiv \sum_{t=1}^{T} W_t f_t. \tag{14}$$

From (13) and (14),

$$\tilde{W} = L'\tilde{C}L, \quad \tilde{C} \equiv (Y'Y)^{-1}Y'\tilde{P}Y(Y'Y)^{-1}. \tag{15}$$

From (15),

$$\tilde{C} = L\tilde{W}L'. \tag{16}$$

From (16), it can be seen that L is obtained by using eigen vectors of $\tilde{C}$. From (15), when $f_t = 1, \ f_l = 0 \ (l \neq t)$, then W_t is obtained by using obtained L and C_t as follows:

$$W_t = L'C_tL, \quad C_t \equiv (Y'Y)^{-1}Y'P_tY(Y'Y)^{-1}, \ t = 1, \cdots, T. \tag{17}$$

When $R < n$ and eigen values $\delta_{R+1}, \cdots, \delta_n$ are close to 0, that is, dimensions $R+1$ to n do not have explanatory power for the given dissimilarity data, (10) can be approximately represented as follows by using lower R dimensions:

$$\overline{P} = V\Delta V' \approx \tilde{V}\tilde{\Delta}\tilde{V}', \tag{18}$$

where $\tilde{V} = (v_{i\lambda}) = (\mathbf{v}_1, \cdots, \mathbf{v}_R), \quad i = 1, \cdots, n, \ \lambda = 1, \cdots, R,$ and $\tilde{\Delta}$ is a diagonal matrix whose diagonal elements are eigen values $\delta_1, \cdots, \delta_R$. Then using $\tilde{V}$ and $\tilde{\Delta}$ in (18), (11) can be approximately represented as follows:

$$\overline{P} \approx \tilde{V}\tilde{\Delta}^{\frac{1}{2}}\tilde{\Delta}^{\frac{1}{2}}\tilde{V}' = (\tilde{V}\tilde{\Delta}^{\frac{1}{2}})(\tilde{V}\tilde{\Delta}^{\frac{1}{2}})', \tag{19}$$

where $\tilde{\Delta}^{\frac{1}{2}}$ is a diagonal matrix whose diagonal elements are eigen values $\sqrt{\delta_1}, \cdots, \sqrt{\delta_R}$. From (19), Y and X shown in (12) and (13) can be estimated as follows:

$$\hat{Y} = \tilde{V}\tilde{\Delta}^{\frac{1}{2}}, \quad \hat{X} = \hat{Y}\hat{L} = \tilde{V}\tilde{\Delta}^{\frac{1}{2}}\hat{L}, \tag{20}$$

where $\hat{X} = (\hat{x}_{i\lambda}), \quad i = 1, \cdots, n, \quad \lambda = 1, \cdots, R,$ and the elements of $\hat{X}$ show values of the coordinate of n objects in R dimensional space when $R < n$. Therefore, the elements of $\hat{X}$ in (20) are the estimate of $x_{i\lambda}$ in the INDSCAL model shown in (4). In order to determine a $R \times R$ orthogonal matrix $\hat{L}$ in (20), the following linear combination is defined, where $f_1, \cdots, f_T$ are any scalars.

$$\tilde{P} = \sum_{t=1}^{T} P_t f_t \approx \hat{\tilde{P}} = \sum_{t=1}^{T} \hat{P}_t f_t = \hat{X}\left(\sum_{t=1}^{T} \hat{W}_t f_t\right)\hat{X}' = \hat{X}\hat{\tilde{W}}\hat{X}', \ \hat{\tilde{W}} \equiv \sum_{t=1}^{T} \hat{W}_t f_t, \quad \hat{P}_t \equiv \hat{X}\hat{W}_t\hat{X}', \tag{21}$$

where $\hat{W}_t$ is a diagonal matrix consisted of elements $w_{t1}, \cdots, w_{tR}$. From (20) and (21),

$$\hat{\tilde{W}} = \hat{L}'\hat{\tilde{C}}\hat{L}, \quad \hat{\tilde{C}} \equiv (\hat{Y}'\hat{Y})^{-1}\hat{Y}'\hat{\tilde{P}}\hat{Y}(\hat{Y}'\hat{Y})^{-1} \approx (\hat{Y}'\hat{Y})^{-1}\hat{Y}'\tilde{P}\hat{Y}(\hat{Y}'\hat{Y})^{-1}. \tag{22}$$

From (22),

$$\hat{\tilde{C}} = \hat{L}\hat{\tilde{W}}\hat{L}'. \tag{23}$$

Therefore, $\hat{L}$ is obtained by using eigen vectors of $\hat{\tilde{C}}$. From (22), when $f_t = 1, \ f_l = 0 \ (l \neq t)$, then $\hat{W}_t$ is obtained by using obtained $\hat{L}$ and $\hat{C}_t$ as follows:

$$\hat{W}_t = \hat{L}'\hat{C}_t\hat{L}, \quad \hat{C}_t \equiv (\hat{Y}'\hat{Y})^{-1}\hat{Y}'\hat{P}_t\hat{Y}(\hat{Y}'\hat{Y})^{-1} \approx (\hat{Y}'\hat{Y})^{-1}\hat{Y}'P_t\hat{Y}(\hat{Y}'\hat{Y})^{-1}, \ t = 1, \cdots, T. \tag{24}$$

So, once we obtain dissimilarity data D_t^2 over T subjects shown in (6), then using the Young-Householder transformation shown in (7), P_t is obtained. Using the P_t and (19), $\hat{Y}$ shown in (20) is obtained. Then using the obtained $\hat{Y}$ and $\tilde{P}$ shown in (14), $\hat{\tilde{C}}$ can be obtained by using (22). And by using the eigenvalue decomposition of $\hat{\tilde{C}}$, we obtain $\hat{L}$ by using R eigen vectors shown in (23), and using the obtained $\hat{Y}$ and $\hat{L}$, we can obtain $\hat{X}$ by using (20). $\hat{W}_t$ is obtained by using P_t, $\hat{Y}$, $\hat{L}$, and (24).

3. Fuzzy Clustering for 3-Way Data

Suppose Z_t be a given data matrix consisted of n objects with respect to p variables at a subject t called a 3-way data and shown as follows:

$$Z_t = (z_{ia,t}), \quad i = 1,\cdots,n, \;\; a = 1,\cdots,p, \;\; t = 1,\cdots,T. \tag{25}$$

In order to obtain the same clusters over the T subjects, the following $nT \times p$ super matrix $\tilde{Z}$ is created.

$$\tilde{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_T \end{pmatrix} = (\tilde{z}_{ja}), \quad j = 1,\cdots,nT, \;\; a = 1,\cdots,p. \tag{26}$$

The purpose of this fuzzy clustering is to classify the nT objects into K clusters. The state of the fuzzy clustering is represented by a partition matrix:

$$\tilde{U} = \begin{pmatrix} U_1 \\ \vdots \\ U_T \end{pmatrix} = (\tilde{u}_{jk}), \quad U_t = (u_{ik,t}), \quad j = 1,\cdots,nT, \;\; i = 1,\cdots,n, \;\; k = 1,\cdots,K, \;\; t = 1,\cdots,T, \tag{27}$$

where $\tilde{u}_{jk}$ is a degree of belongingness of an object j which is shown as $\tilde{\mathbf{z}}_j = (\tilde{z}_{j1},\cdots,\tilde{z}_{jp})$ to a fuzzy cluster k and $u_{ik,t}$ is a degree of belongingness of an object i to the same fuzzy cluster k at a subject t. From (27), it can be seen that the obtained K fuzzy clusters are the same over T subjects. In general, $\tilde{u}_{jk}$ satisfies the following conditions:

$$\tilde{u}_{jk} \in [0,1], \quad \sum_{k=1}^{K} \tilde{u}_{jk} = 1, \quad j = 1,\cdots,n. \tag{28}$$

From these conditions in (28), it can be seen that the result of fuzzy clustering has a compositional data structure. Fuzzy c-means (FCM)[7] is one of the methods of fuzzy clustering. FCM is a method which minimizes the weighted within-class sum of squares:

$$J(\tilde{U},\mathbf{g}_1,\cdots,\mathbf{g}_K) = \sum_{j=1}^{n}\sum_{k=1}^{K} \tilde{u}_{jk}^{m} d^2(\tilde{\mathbf{z}}_j,\mathbf{g}_k), \tag{29}$$

where $\mathbf{g}_k = (g_{k1},\cdots,g_{kp})$ denotes the values of the centroid of a cluster k, $d^2(\tilde{\mathbf{z}}_j,\mathbf{g}_k)$ is the squared Euclidean distance between $\tilde{\mathbf{z}}_j$ and $\mathbf{g}_k$. The exponent m that determines the degree of fuzziness of the clustering is chosen from $(1,\infty)$ in advance. By minimizing (29) under the conditions shown in (28), we obtain the solutions $\tilde{U},\mathbf{g}_1,\cdots,\mathbf{g}_K$.

4. Individual Compositional Cluster Analysis

Suppose a 3-way data $Z_1,\cdots,Z_T$ shown in (25) is given. Then the following squared Euclidean distance can be caluculated.

$$d_{ij,t}^2 = \sum_{a=1}^{p} (z_{ia,t} - z_{ja,t})^2, \quad i,j = 1,\cdots,n, \;\; t = 1,\cdots,T. \tag{30}$$

Then using this as a given 3-way dissimilarity data $D_t^2 = (d_{ij,t}^2)$ in (6), and by using INDSCAL model, we can obtain the coordinate values for all objects over all subjects in a fewer dimensional space which is $\hat{X}$ shown in (20), as well as the weights of individual subjects for the obtained dimensions which is $\hat{W}_t$ shown in (24). However, if the number of objects, n, is very large, then large information of data ($n - R$ dimensions data information) has to be deleted in order to obtain the result of coordinate values in the R dimensional space when R is a much smaller number (fewer number) of dimensions than n ($R << n$). In this case, obviously it is very difficult to fit the observed data shown in (30) to the INDSCAL model, and number of dimensions has to be large. In order to solve this problem, we utilize the fact in which fuzzy clustering result, $U^t, t = 1,\cdots,T$, shown in (27) are obtained with respect to the same fuzzy clusters over T subjects and the number of clusters, K, is much smaller than the number of objects, n. First, we define the following transformation from the coordinate values of n objects shown by $\hat{X}$ shown in (20) to weighted coordinate values of K clusters shown as $\tilde{X}_t$ as follows.

$$\tilde{X}_t = U_t^{'}\hat{X}, \quad t = 1,\cdots T. \tag{31}$$

Where $\hat{X}$ is a $n \times R$ matrix which is consisted of common coordinate values over T subjects and this is a result of INDSCAL shown in (20). $U_t, t = 1,\cdots T$ show $n \times K$ matrixes obtained as a fuzzy clustering result shown in (27). Therefore, (31) means that we combine a result of fuzzy clustering for 3-way data and a result of INDSCAL. By using the combination, from (31), we can see that $\tilde{X}_t, t = 1,\cdots,T$ are $K \times R$ matrixes, where K is the number of fuzzy clusters and these fuzzy clusters are guaranteed to be the same over the T subjects, because these fuzzy clusters are obtained as the result of the fuzzy clustering for 3-way data described in section 3. In addition, R dimensions are obtained as a result of INDSCAL. Therefore, since $\tilde{X}_t, t = 1,\cdots,T$ are consisted of the "same" K fuzzy clusters and the "same" R dimensions over the T subjects, $\tilde{X}_t, t = 1,\cdots T$ can be regarded as a 3-way data. In addition, by using the 3-way data $\tilde{X}_t, t = 1,\cdots,T$, replaced by original 3-way data $Z_t, t = 1,\cdots,T$, we can reduce the number of objects from n to K, which can overcome the problem of large number of objects and the difficulty to fit the large amount of data to the lower dimensional space. The squared Euclidean distance of $\tilde{X}_t, t = 1,\cdots T$ can be calculated as follows.

$$\tilde{d}_{kl,t}^2 = \sum_{\lambda=1}^{R}(\tilde{x}_{k\lambda,t} - \tilde{x}_{l\lambda,t})^2, \quad k,l = 1,\cdots,K, \ t = 1,\cdots,T, \tag{32}$$

where $\tilde{X}_t = (\tilde{x}_{k\lambda,t}), \quad k = 1,\cdots,K, \quad \lambda = 1,\cdots,R, \quad t = 1,\cdots,T.$ By applying the 3-way dissimilarity data shown in (32) to the INDSCAL model, the solution

$$\hat{X} = (\hat{x}_{k\lambda}), \quad k = 1,\cdots,K, \ \lambda = 1,\cdots,\tilde{R} \tag{33}$$

shown in (20) which is consisted of common coordinate values of K clusters in $\tilde{R}$ dimensional space where $\tilde{R}$ could be much smaller than R, since $n >> K > \tilde{R}$. Also, the solution $\hat{W}_t = (\hat{w}_{t\lambda}), t = 1,\cdots,T, \lambda = 1,\cdots,\tilde{R}$ shown in (24) are obtained to show the difference of T subjects. From (5), the estimated individual coordinate values $\hat{o}_{k\lambda,t}, t = 1,\cdots,T$ are obtained as follows:

$$\hat{O} = (\hat{o}_{k\lambda,t}) = (\sqrt{\hat{w}_{t\lambda}}\hat{x}_{k\lambda}), \quad k = 1,\cdots,K, \ \lambda = 1,\cdots,\tilde{R}, \ t = 1,\cdots,T. \tag{34}$$

The solutions $\hat{X}$ and $\hat{O}$ shown in (33) and (34) can be jointly plotted in the same space spanned by $\tilde{R}$ dimensions, since the $\tilde{R}$ dimensions are exactly the same for both solutions $\hat{X}$ and $\hat{O}$ mathematically. Then we can compare the location of K clusters as well as the clusters including the individual difference visually in the $\tilde{R}$ dimensional vector space.

5. Selection of Adaptable Number of Clusters

We propose the following criterion which determines an adaptable number of clusters with the cluster centroids $\mathbf{g}_k = (g_{k1},\cdots,g_{kp}), k = 1,\cdots,K$ shown in (29).

$$H(K) = (\sum_{k=1}^{K}\sum_{a=1}^{45}(g_{ka} - \bar{g}_a)^2)/(\sum_{k=1}^{K}(s_k - \bar{s}_k)^2), \quad \bar{g}_a = \frac{1}{K}\sum_{k=1}^{K}g_{ka}, \quad s_k = \tilde{s}_k/(\min_{1 \le k \le K}\tilde{s}_k), \quad \bar{s}_k = \frac{1}{K}\sum_{k=1}^{K}s_k, \tag{35}$$

where $\tilde{s}_k$ shows number of objects in a cluster k. The numerator of the criterion $H(K)$ shows variance of the values of obtained cluster centroids over the variables. Therefore, the fact that the value of the numerator is large means the obtained cluster centers are separated well and should be determined as an adaptable number of clusters when $H(K)$ has a maximum value. However, in this case, the criterion will be very sensitive to the outlier, that is, if there is an extremely separated singleton cluster, then the value of criterion becomes large and mistakenly suggest a wrong number of clusters. In order to avoid this problem, we add the denominator part which shows the variance of cluster size ratio with respect to the minimum size of the cluster for the purpose of avoiding the situation of outlier cluster.

6. Numerical Example

We use a dataset of sensor data of daily and sports activities performed by 8 subjects with respect to 45 variables[8,9]. The dataset consists of 19 regular daily and sports activities, and we select the activity of running on a treadmill with a speed of 8 km/h. For this activity, we use data which is consisted of 1250 times (or objects), 45 variables, and 8 subjects. For times (or objects), we select the first 50 seconds of data whose number of times is 1250. For the variables, the data was observed for 45 (5 body positions $\times$ 9 kinds of sensors) variables which consists of 5 body-worn sensor units positioned on the torso (T), right arm (RA), left arm (LA), right leg (RL), left leg (LL), and each sensor unit has 9 kinds of information including x-axial accelerometers (xacc), y-axial accelerometers (yacc), z-axial accelerometers (zacc), x-axial gyroscopes (xgyro), y-axial gyroscopes (ygyro), z-axial gyroscopes (zgyro), x-axial magnetometers (xmag), y-axial magnetometers (ymag), z-axial magnetometers (zmag). For the subjects, eights subjects (4 females, 4 males) are performed for this activity. We treat this data as a 3-way data shown in (25), where n is 1250, p is 45, and T is 8. Then a super matrix $\tilde{Z}$ shown in (26), whose number of rows is 10000 (=1250$\times$ 8) and the number of columns is 45, is created. By applying this data to the fuzzy clustering for 3-way data shown in (29) under the conditions in (28), the result of the clustering shown in (27) is obtained. Fig. 1 shows the result of the values of criterion shown in (35). The abscissa shows the number of clusters and ordinate shows the values of the criterion, $H(K)$. From this result, we select the number of clusters as 3. Fig. 2 shows the result of fuzzy clustering for 3-way data shown in (27). In this figure, the abscissa shows the number of times, so from 1 to 1250 shows a result of the first subject, from 1251 to 2500 is for subject 2, from 2501 to 3750 is for subject 3, from 3751 to 5000 is for subject 4, from 5001 to 6250 is for subject 5, from 6251 to 7500 is for subject 6, from 7501 to 8750 is for subject 7, and from 8751 to 10000 is for the subject 8.

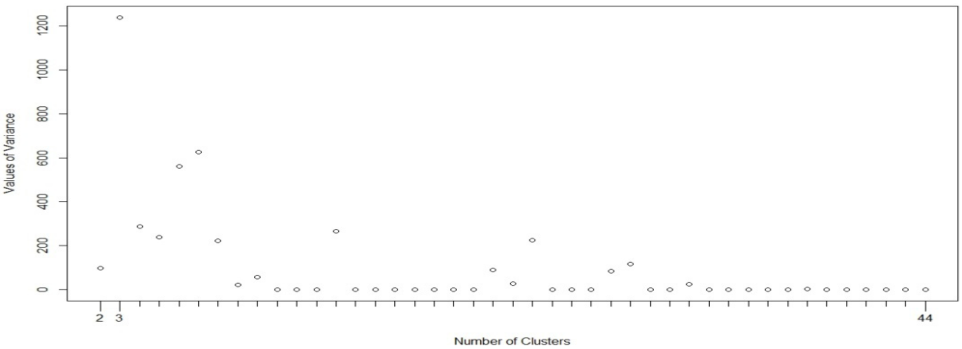

Fig. 1. Criterion values for selecting number of clusters

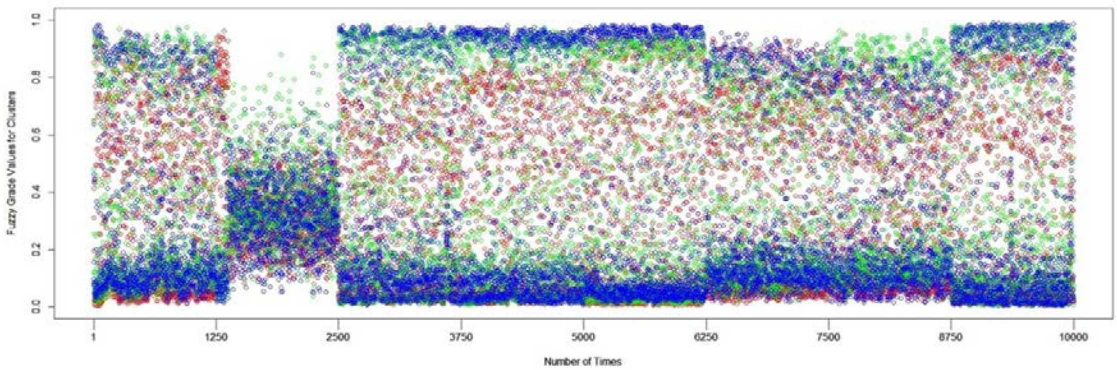

Fig. 2. Result of fuzzy clustering for 3-way data

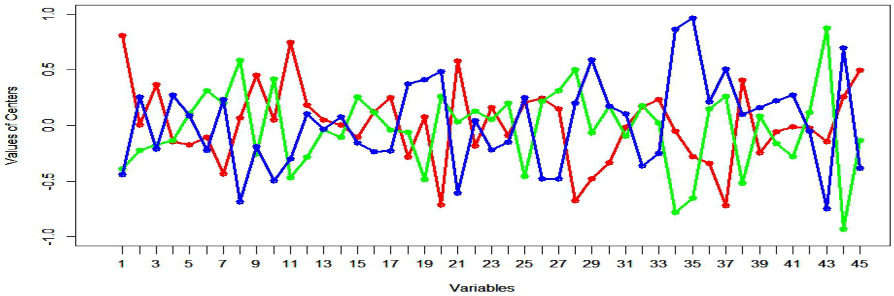

Fig. 3. Result of centroids of fuzzy clusters

45 variables are as follows: 1: T_xacc, 2: T_yacc, 3: T_zacc, 4: T_xgyro, 5: T_ygyro, 6: T_zgyro, 7: T_xmag, 8: T_ymag, 9: T_zmag, 10: RA_xacc, 11: RA_yacc, 12: RA_zacc, 13: RA_xgyro, 14: RA_ygyro, 15: RA_zgyro, 16: RA_xmag, 17: RA_ymag, 18: RA_zmag, 19: LA_xacc, 20: LA_yacc, 21: LA_zacc, 22: LA_xgyro, 23: LA_ygyro, 24: LA_zgyro, 25: LA_xmag, 26: LA_ymag, 27: LA_zmag, 28: RL_xacc, 29: RL_yacc, 30: RL_zacc, 31: RL_xgyro, 32: RL_ygyro, 33: RL_zgyro, 34: RL_xmag, 35: RL_ymag, 36: RL_zmag, 37: LL_xacc, 38: LL_yacc, 39: LL_zacc, 40: LL_xgyro, 41: LL_ygyro, 42: LL_zgyro, 43: LL_xmag, 44: LL_ymag, 45: LL_zmag

The first cluster is shown as a red color, the second cluster is represented as green, and the third cluster is shown as blue. Since the subjects 1, 2, 6, 7 are females and subjects 3, 4, 5, 8 are males, from this result, we can see the tendency of distinguishing between female and male. Fig. 3 shows the values of obtained centroids of fuzzy clusters. The red line shows the first cluster, green shows the second cluster, and the blue shows the third cluster. In this figure, the abscissa shows the number of variables and the ordinate shows values of centroids. From this result, it can be seen that the centroids of three clusters are quite different from each other that is, the obtained three clusters are separated well. Using the 3-way data, the squared Euclidean distance $D_t^2 = (d_{ij,t}^2),\ t = 1,\cdots,T$ shown in (30) is calculated. Then the calculated distance is applied to the INDSCAL model. As the first step, the 3-way distances are transformed to $P_t, t = 1,\cdots,8$ shown in (7) and calculates $\overline{P}$ shown in (9). By the eigen value decomposition of $\overline{P}$, eigen values $\delta_1,\cdots,\delta_{1250}$ are obtained. Fig. 4. Shows the eigen values of 1250 dimensions. We selected 80 eigen values, since the cumulative proportion of the 80 eigen values is almost 0.8, $\sum_{i=1}^{80}\delta_i / \sum_{i=1}^{1250}\delta_i = 0.81$. However, the 80 dimensions are large to capture the data structure. Therefore, we transform the data to weighted data $\tilde{X}_t, t = 1,\cdots 8$ by using the fuzzy clustering result shown in Fig. 2 and the transformation shown in (31). By applying the squared Euclidean distance of $\tilde{X}_t, t = 1,\cdots 8$ shown in (32) to the INDSCAL model, we obtain the solution of three clusters which are common coordinate values of the three clusters over 8 subjects in two dimensional space shown in (33), and the solution of the three clusters with individual difference of the 8 subjects included which is shown in (34). Fig. 5 shows the results of common clusters shown in (33) and clusters including an individual subject's difference shown in (34). In this figure, "C1", "C2", and "C3" show the three common clusters and "P1C1" – "P8C1" are the first to eighth subject's difference including the first cluster's location. "P1C2" – "P8C2" and "P1C3" – "P8C3" are the location points of individual difference included cluster 2 and cluster 3. The cumulative proportion of two eigen values for the two dimensions is almost 1.0. This means the result shown in Fig. 5 can almost perfectly explain the data structure shown in (31) and the proposed method can successfully reduce the number of dimensions from 1250 to 2 dimensions. From Fig. 5, it can be seen that dimension 1 shows the difference between clusters 2 and 3, and dimension 2 shows the difference of cluster 1 from other two clusters. Figs. 6 - 8, show the magnified figures of Fig. 5 at areas of the neighborhoods of "C1", "C2", and "C3", respectively. From these figures, we can see that subjects 3, 4, and 7 have a similar tendency of the commonly obtained cluster feature, and other subjects have stronger individuality. Fig. 9 shows the difference of weights obtained from the diagonal elements $\hat{W}_t,\ t = 1,\cdots,8$ shown in (24). The result of $\hat{W}_t,\ t = 1,\cdots,8$ is not always obtained as the diagonal matrixes, however, in this time, the values of the non-diagonal part are relatively small enough when compared with the values of the diagonal part, so we only capture the diagonal part. In Fig. 9, the abscissa shows the values of weights of dimension 1 and the ordinate shows the values of weights for dimension 2. From this result, we can see the difference of the 8 subjects, for example, it can be seen that subjects 1 and 2 are more strongly related with dimension 2 when compared with the reliance to the first dimension. On the other hand, subjects 5, 6, and 8 are more reliant with dimension 1 than the second dimension. Subjects 3, 4, and 7 have balanced reliance, that is, for the two dimensions evenly related to each other. We can see that dimension 1 seems to show the Leg movements such as RL_xmag, RL_ymag, LL_xmag, LL_ymag, since Fig. 5 shows that dimension 1 shows the difference between clusters 2 and 3, and from Fig. 3, variables of Leg movements has big difference between clusters 2 and 3. Dimension 2 is most likely to show the relation of accelerometers, since from Fig. 5, dimension 2 shows difference of cluster 1 from other two clusters, as well as

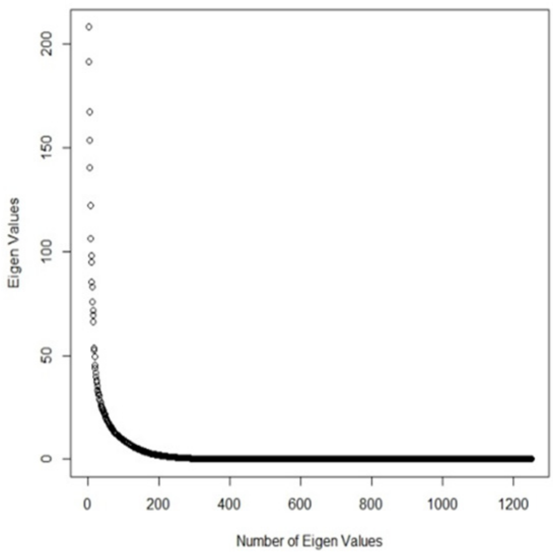

Fig. 4. Eigen values

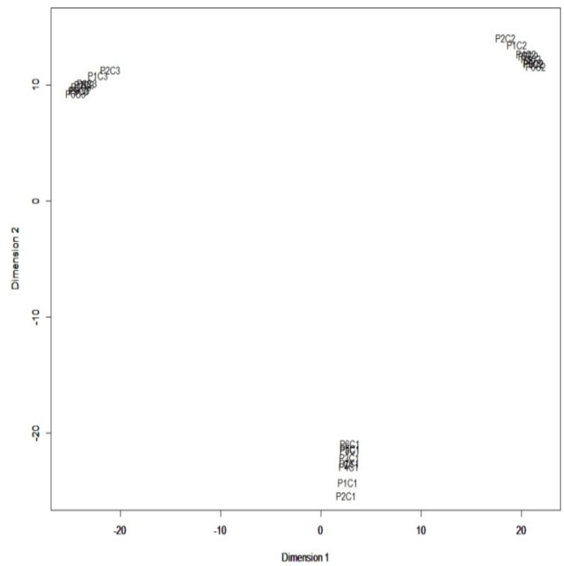

Fig. 5. Result of proposed Individual Compositional Cluster Analysis

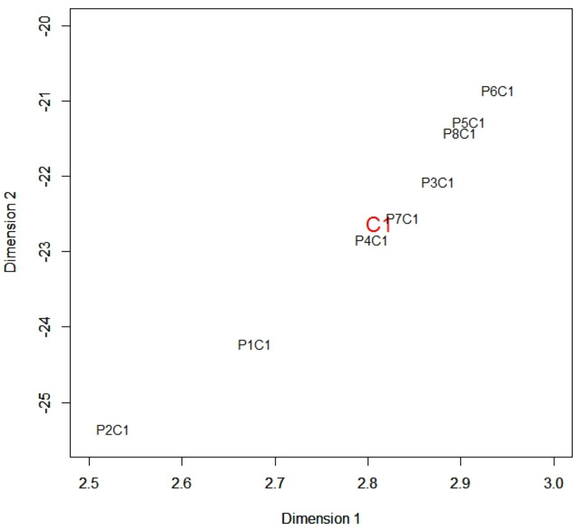

Fig. 6. Subjects around cluster 1 in area magnified Fig. 5

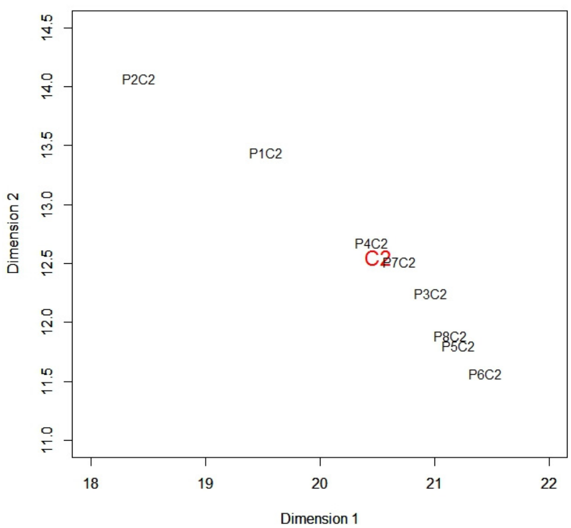

Fig. 7. Subjects around cluster 2 in area magnified Fig. 5

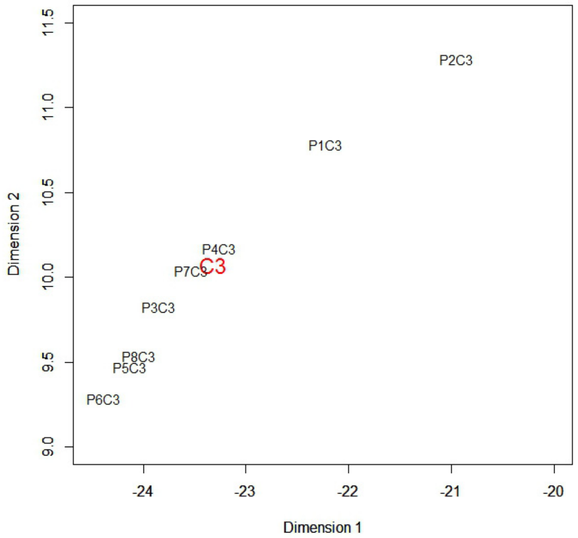

Fig. 8. Subjects around cluster 3 in area magnified Fig. 5

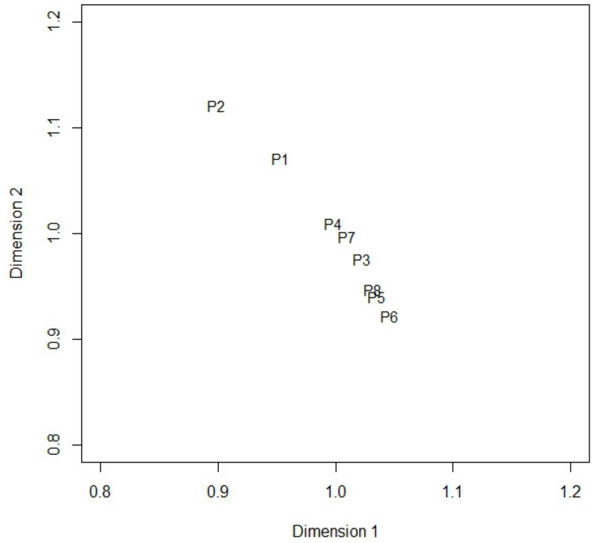

Fig. 9. Result of weights

Fig. 3 shows that the value of cluster 1 tends to differ from the values of clusters 2 and 3 at the variables related with accelerometers. From figures 5-8, we can see the difference of 8 subjects visually in two dimensional space including clusters capturing the latent features of the original 3-way data consisted of 1250 times (or objects), 45 variables, and 8 subjects.

7. Conclusions

With the observed data consisting of objects, variables, and subjects, this paper proposes a method named Individual Compositional Cluster Analysis for showing the individual difference of subjects visually in a lower dimensional Euclidean vector space including clustering of objects. This method is especially useful when the number of objects is large, due to the number of clusters being much smaller than the number of objects. Although the reduction of the number of dimensions tends to lose a lot of information contained in the original data, with this method, the loss of information is reduced due to the inclusion of clustering in the data reduction. In addition, for the proposed method, a fuzzy clustering for 3-way data and a criterion to select an adaptable number of clusters are proposed. Several numerical examples using 3-way time series data show a better performance of the proposed method. Considering the continuous change of the time, the extension of the proposed method by using the higher-order metrics is a future problem.

References

1. Schonemann PH, An algebraic solution for class of subjective metrics models. *Psychometrika* 1972; 37: 441-451.
2. Cox TF, Cox MAA. *Multidimensional scaling.* Chapman & Hall; 2001.
3. Sato-Ilic M, Multidimensional joint scale and cluster analysis. *Procedia Computer Science, Elsevier* 2015; 61:11-17.
4. Gower JC, Some distance properties of latent roots and vector methods used in multivariate analysis. *Biometrika* 1966; 53:325-338.
5. Torgerson WS, Theory and methods of scaling. Wiley; 1958.
6. Young G, Householder AS. Discussion of a set of points in terms of their mutual distances. *Psychometrika* 1938; 3:19-22.
7. Bezdek JC, *Pattern recognition with fuzzy objective function algorithms*. Plenum; 1981.
8. Altun K, Barshan B. Human activity recognition using inertial/magnetic sensor units. In: Salah AA, Gevers T, Sebe N, Vinciarelli A, editors. *HBU 2010*. LNCS 6219, Springer; Berlin, Heidelberg, 2010. p. 38-51.
9. UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/index.html